

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2005, 17:16:07 ; Search time 162 Seconds
(without alignments)
128.920 Million cell updates/sec

Title: US-10-644-927-1

Perfect score: 304

Sequence: 1 KTYGTNGVHCTKNSLWGVK.....GRLQDILLGWATGAFGKTEH 54

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_16Dec04:*

- 1: Geneseqp1980s:*
- 2: Geneseqp1990s:*
- 3: Geneseqp2000s:*
- 4: Geneseqp2001s:*
- 5: Geneseqp2002s:*
- 6: Geneseqp2003as:*
- 7: Geneseqp2003bs:*
- 8: Geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69.5	22.9	420	6	ADA33974 Acinetoba
2	64	21.1	61	2	Aw17990 Sakacin P
3	62.5	20.6	558	8	ADR89438 crybun2-o
4	62	20.4	42	2	AAW11633 Lactobaci
5	60	19.7	37	2	AAW66446 Catonic
6	60	19.7	37	3	AAW91745 Catonic
7	60	19.7	300	2	AAW56481 CD38. 3/2
8	60	19.7	300	2	AAW37928 Amino aci
9	60	19.7	300	4	AAW69069 Human CD3
10	60	19.7	300	5	ABG61817 Prostata
11	60	19.7	300	6	ABU09711 CD38 prot
12	60	19.7	300	6	ABU09714 CD38 prot
13	60	19.7	300	7	ADD18904 Human dis
14	60	19.7	300	7	ADD45262 Human pro
15	60	19.7	300	7	ADN40030 Cancer/an
16	60	19.7	300	8	ADM32177 Human CD3
17	60	19.7	300	8	ADQ59526 Human can
18	60	19.7	300	8	ADQ88192 Human 903
19	60	19.7	688	4	ABG06883 Novel hum
20	59.5	19.6	513	2	AAW79755 Euphorbia
21	59	19.4	44	2	AAR14564 Bacterioc
22	59	19.4	44	2	AAR91745 Brevicin,
23	59	19.4	62	2	AAR25078 PA-1 bact
24	59	19.4	300	6	ABU09712 CD38 prot
25	59	19.4	399	8	ADN23488 Bacterial

ALIGNMENTS

RESULT 1

ADA33974
ID ADA33974 standard; protein; 420 AA.

XX AC ADA33974;

XX XX 20-NOV-2003 (first entry)

XX XX Acinetobacter baumannii protein #1135.

DE Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;

XX KW Acinetobacter baumannii; bacterial disease; antibacterial; vaccine;

XX KW plant biocontrol agent.

XX OS Acinetobacter baumannii.

XX XX US6562958-B1.

XX PD 13-MAY-2003.

XX PF 04-JUN-1999; 99US-00328352.

XX PR 09-JUN-1998; 98US-0088701P.

XX (GENO-) GENOME THERAPEUTICS CORP.

XX PA Breton G, Bush D;

XX DR WPI; 2003-576092/54.

XX N-PSDB; ADA29848.

XX PT New Acinetobacter baumannii proteins and nucleic acids, useful as reagents for diagnosing a bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, or as biocontrol agents for plants.

XX PS Example; SEQ ID NO 5261; 328pp; English.

XX CC The invention relates to isolated Acinetobacter baumannii nucleic acids. The A. baumannii nucleic acids and polypeptides are useful as reagents for diagnosing a bacterial disease, as components of antibacterial vaccines, as targets for antibacterial drugs, to detect the presence of A. baumannii and other Acinetobacter species in a sample, in screening compounds for the ability to interfere with the A. baumannii life cycle or to inhibit A. baumannii infection, and as biocontrol agents for plants. The present sequence represents the amino acid sequence of an A. baumannii protein.

XX CC Sequence 420 AA;

Aab50093 GB1 prote
Adc64559 Trichodes
Adq07360 Trichodes
Abb71088 Drosophil
Abu44608 Protein e
Ame69065 Phototrab
Adb05473 Enterocin
Adf04643 Bacterial
Adc95053 E. faeciu
Abu42656 Protein e
Aap39314 Staphyloc
Adb04851 S. epiderm
Adb04803 Staphyloc
Aay06359 Fusarium
Aay84337 Amino aci
Aab14872 Fusarium
Aau77580 F. javani
Aau77424 Fusarium
Add46266 Rat Prote
Ade62845 Rat Prote

26 59 19.4 753 4 AAB50093
27 57.5 18.9 461 7 ADC64559
28 57.5 18.9 461 8 ADQ07360
29 57.5 18.9 643 4 ABB71088
30 57 18.8 581 6 ABU44608
31 57 18.8 1565 6 AME69065
32 56.5 18.6 45 5 ADB05473
33 56.5 18.6 242 7 ADF04643
34 56 18.4 66 7 ADC95053
35 56 18.4 6641 6 ABU42656
36 56 18.4 10182 5 AAP39314
37 56 18.4 10203 8 ADB04851
38 56 18.4 10203 8 ADB04803
39 55.5 18.3 244 2 AAY06359
40 55.5 18.3 244 3 AAY84337
41 55.5 18.3 244 3 AAB14872
42 55.5 18.3 244 5 AAU77580
43 55.5 18.3 244 5 AAU77424
44 55.5 18.3 1382 7 ADD46266
45 55.5 18.3 1382 7 ADE62845

RESULT 5
AAW66446
ID AAW6
XX

```

OS Unidentified.
XX WO9965506-A2.
XX PD 23-DEC-1999.
XX PF 14-JUN-1999; 99WO-CA000552.
XX PR 12-JUN-1998; 98US-00096541.
XX PA (MTCR-) MICROLOGIX BIOTECH INC.
XX PI Friedland HD, Krieger TJ, Taylor R, Erfile D, Fraser JR, West MHP;
XX WPI; 2000-223549/19.
XX Novel pharmaceutical composition containing optionally activated
PT polyoxyalkylene-modified cationic peptides, useful for treating tumors.
XX Disclosure; Page 11; 94pp; English.
XX This sequence represents a cationic peptide amino acid sequence, which
CC can be used in the pharmaceutical composition of the invention. The
CC invention relates to a pharmaceutical composition containing at least one
CC activated polyoxyalkylene (APO)-modified cationic peptide. The
CC modification of peptides with APO increases their activity against tumour
CC cells, including those with a multidrug resistant phenotype. The
CC pharmaceutical composition can be used to treat tumours, specifically
CC lymphoma, leukaemia, multiple myeloma, or tumours of breast, lung, ovary,
CC cervix, uterus, skin, prostate, liver and colon
XX Sequence 37 AA;
SQ
Query Match 19.7%; Score 60; DB 3; Length 37;
Best Local Similarity 48.4%; Pred. No. 3.5;
Matches 15; Conservative 2; Mismatches 4; Indels 10; Gaps 3;
QY 3 YYGTNGVHCTKNSL---WGKV-----RLKN 24
DB 2 YYG-NGVHCTKSGSVNMGAFSAGVHRLAN 31
RESULT 7
-AAR56481
ID AAR56481 standard; protein; 300 AA.
XX AC AAR56481;
XX 25-MAR-2003 (revised)
XX 22-FEB-1995 (first entry)
XX CD38.
XX CD38; therapeutic; diagnostic.
XX Homo sapiens.
XX WO9417184-A1.
XX 04-AUG-1994.
XX 27-JAN-1994; 94WO-US000517.
XX 29-JAN-1993; 93US-00010905.
XX (SCHE ) SCHERING CORP.
XX (STRD ) UNIV LELAND STANFORD JUNIOR.
XX Parkhouse RME, Santos-Argumedo L, Grimaldi JC, Bazan JF, Heath A;
XX Howard MC, Goodnow CC;
XX WPI; 1994-264098/32.
XX N-PSDB; AAO70644.
XX Modulation of responses of lymphocytes - using antibody to CD38, a
PT soluble CD38 fragment or a modulator of ADP-ribosyl cyclase or cyclic ADP
PT -ribosyl hydrolase.
XX Disclosure; Page 46; 54pp; English.
XX The protein is a human CD38 molecule which is used to modulate the
CC physiological response of a lymphocyte, which is useful for the
CC stimulation/inhibition of lymphocyte growth or differentiation,
CC particularly for the establishment of antigen tolerance. (Updated on 25-
CC MAR-2003 to correct PN field.)
XX Sequence 300 AA;
SQ
Query Match 19.7%; Score 60; DB 2; Length 300;
Best Local Similarity 35.0%; Pred. No. 43;
Matches 14; Conservative 7; Mismatches 19; Indels 0; Gaps 0;
QY 5 GTNGVHCTKNSLWKGKVLKMKYDQNTTYNGRLQDILLGW 44
DB 113 GTQTPCNCILLWSRIKDLAHQFTQVQORDMFTLEDLLGY 152
RESULT 8
AAW37928
ID AAW37928 standard; protein; 300 AA.
XX AC AAW37928;
XX 09-SEP-1998 (first entry)
XX Amino acid sequence of CD38.
XX CD38; CD38 peptide fragment; anti-CD38 autoantibody; autoimmune disease;
XX insulin dependent diabetes mellitus; type II diabetes.
XX Mammalia.
XX WO9816245-A1.
XX 23-APR-1998.
XX 11-APR-1997; 97WO-JP001259.
XX 15-OCT-1996; 96JP-00272537.
XX (SHIO ) SHIONOGI & CO LTD.
XX Taminato T;
XX WPI; 1998-251054/22.
XX N-PSDB; AAV29155.
XX Assaying anti-CD38 auto-antibody - useful for detecting auto-immune
PT disease, e.g. type II diabetes.
XX Disclosure; Page 8-10; 17pp; Japanese.
XX This is the amino acid sequence of the CD38 protein, of which a fragment
CC is used in the method of the invention to create the anti-CD38
CC autoantibody. The method is used for detecting autoimmune disease, e.g.
CC insulin dependent diabetes mellitus or type II diabetes
XX Sequence 300 AA;
SQ
Query Match 19.7%; Score 60; DB 2; Length 300;
Best Local Similarity 35.0%; Pred. No. 43;
Matches 14; Conservative 7; Mismatches 19; Indels 0; Gaps 0;
QY 5 GTNGVHCTKNSLWKGKVLKMKYDQNTTYNGRLQDILLGW 44
DB 113 GTQTPCNCILLWSRIKDLAHQFTQVQORDMFTLEDLLGY 152

```

```
RESULT 9
AAB69069
ID AAB69069 standard; protein; 300 AA.
XX
AC AAB69069;
XX
XX 19-APR-2001 (first entry)
XX
DE Human CD38 protein sequence SEQ ID NO:2.
XX
XX Human; CD38; diabetes mellitus; detection; cyclic ADP-ribose; CADPR.
XX
XX Homo sapiens.
XX
PN JP2000316578-A.
XX
PD 21-NOV-2000.
XX
XX 12-MAY-1999; 99JP-00131955.
XX
PR 12-MAY-1999; 99JP-00131955.
XX
XX (BMLB-) BML KK.
XX
XX (KANE/) KANETSUKA A.
XX
XX (OKAM/) OKAMOTO H.
XX
DR WPI; 2001-128255/14.
DR N-PSDB; AAF32487.
XX
XX Detecting onset of diabetes mellitus comprises detecting specific gene
XX mutations in the CD38 gene.
XX
XX Example; Page 11-12; 19pp; Japanese.
XX
XX The present invention describes a method using a mutation in the CD38
XX gene (involved in the production of cyclic ADP-ribose (CADPR)), to detect
XX the onset of diabetes mellitus. The method is useful for detecting the
XX onset of diabetes mellitus. The present sequence represents human CD38,
XX which is used in an example from the present invention
XX
SQ Sequence 300 AA;
Query Match 19.7%; Score 60; DB 4; Length 300;
Best Local Similarity 35.0%; Pred. No. 43;
Matches 14; Conservative 7; Mismatches 19; Indels 0; Gaps 0;
QY 5 GTNGVHCTKNSLWGVRLKNMKYDQNTTYNGRLQDILLGW 44
DB 113 GTQVPCNKILLWSRIKDLAHQFTQVQORDMFTLEDTLGGY 152

RESULT 10
ABG61817
ID ABG61817 standard; protein; 300 AA.
XX
AC ABG61817;
XX
XX 15-AUG-2002 (first entry)
XX
DE Prostate cancer-associated protein #18.
XX
XX Prostate cancer; prostate tumour tissue; human; mammal; cytostatic.
XX
XX Mammalia.
XX
PN WO200230268-A2.
XX
PD 18-APR-2002.
XX
XX 12-OCT-2001; 2001WO-US032045.
XX
```

```
PR 13-OCT-2000; 2000US-00687576.
PR 08-DEC-2000; 2000US-00733288.
PR 08-DEC-2000; 2000US-00733742.
PR 24-JAN-2001; 2001US-0263957P.
PR 16-MAR-2001; 2001US-0276791P.
PR 16-MAR-2001; 2001US-0276888P.
PR 06-APR-2001; 2001US-0281322P.
PR 24-APR-2001; 2001US-0286214P.
PR 30-APR-2001; 2001US-00847046.
PR 04-MAY-2001; 2001US-0288589P.
XX
XX (SOSB-) EOS BIOTECHNOLOGY INC.
XX
XX Gish KC, Mack DH, Wilson KE, Afar D, Hevezi P;
XX WPI; 2002-471335/50.
XX N-PSDB; ABK92132.
XX
XX Detecting a prostate cancer-associated transcript in a cell in a patient,
XX useful for diagnosing prostate cancer (PC) or screening modulators of PC,
XX by determining if prostate cancer-associated genes are expressed in a
XX prostate tissue.
XX
XX Claim 27; Page 314; 436pp; English.
XX
XX The present invention relates to methods of detecting a prostate cancer-
XX associated transcript in a cell from a patient. The method comprises
XX contacting a biological sample from the patient with prostate cancer-
XX associated polynucleotides (designated PC genes) that selectively
XX hybridise to a sequence that is at least 80% identical to them. The
XX prostate cancer-associated polynucleotide sequences are differentially
XX expressed in prostate tumour tissue or in prostate cancer and are derived
XX from the tissues of various organisms such as humans or other mammals
XX (e.g. mice, sheep and dogs). The methods of the invention are useful for
XX diagnosing and treating prostate cancer in mammals. The prostate cancer-
XX associated genes are useful for diagnosing or treating prostate cancer,
XX as well as for identifying modulators of prostate cancer or agents that
XX inhibit prostate cancer. The nucleic acid sequences are particularly
XX useful in gene therapy, as a vaccine or in antisense applications.
XX ABG61800-ABG61944 represent prostate cancer-associated proteins
XX
SQ Sequence 300 AA;
Query Match 19.7%; Score 60; DB 5; Length 300;
Best Local Similarity 35.0%; Pred. No. 43;
Matches 14; Conservative 7; Mismatches 19; Indels 0; Gaps 0;
QY 5 GTNGVHCTKNSLWGVRLKNMKYDQNTTYNGRLQDILLGW 44
DB 113 GTQVPCNKILLWSRIKDLAHQFTQVQORDMFTLEDTLGGY 152

RESULT 11
ABU09711
ID ABU09711 standard; protein; 300 AA.
XX
XX ABU09711;
XX
XX 03-JUL-2003 (first entry)
XX
DE CD38 protein.
XX
XX CD38; diabetic onset; diabetes; lymphocyte surface marker.
XX
XX Hominidae.
XX
XX US2003027134-A1.
XX
XX 06-FEB-2003.
XX
XX 31-JAN-2001; 2001US-00773307.
XX
XX 31-JAN-2001; 2001US-00773307.
XX
```



```

PR 12-APR-2002; 2002US-0372246P.
PR 05-JUN-2002; 2002US-0386614P.
PR 16-JUL-2002; 2002US-0396839P.
PR 22-JUL-2002; 2002US-0397775P.
PR 23-JUL-2002; 2002US-0397845P.
PR 09-SEP-2002; 2002US-0409450P.
XX
PA (POSB-) EOS BIOTECHNOLOGY INC.
XX
XX Afar D, Aziz N, Ginsburg WM, Gish KC, Glynn R, Hevezi PA;
PI Mack DH, Murray R, Watson SR, Wilson KE, Zlotnik A;
XX
XX WPI; 2003-469649/44.
DR N-PSDB; ADN39813.
XX
XX Determining the presence or absence of a pathological cell in a patient,
PT useful for diagnosing, prognosing or treating cancer, comprises detecting
PT a nucleic acid in a biological sample.
XX
XX Claim 12; SEQ ID NO C400; 1385pp; English.
XX
XX The invention relates to nucleic acids and proteins (ADN38693-ADN40064)
CC whose expression is upregulated or downregulated in specific cancers or
CC other diseases such as angiogenic or fibrotic disorders, and to methods
CC of determining the presence or absence of a pathological cell in a
CC patient by detecting a nucleic acid at least 80% identical to those of
CC the invention or by detecting a polypeptide of the invention. The
CC invention also relates to expression vectors and host cells comprising a
CC nucleic acid of the invention; antibodies which specifically bind a
CC polypeptide of the invention; use of such antibodies for drug targeting;
CC and methods of screening for modulators of activity or expression of the
CC polypeptides and nucleic acids. The nucleic acids, polypeptides,
CC antibodies and methods are useful for diagnosing, prognosing and treating
CC cancer and other conditions such as psoriasis, ischaemia, heart disease,
CC atherosclerosis, inflammatory diseases, autoimmune diseases, retinal
CC neovascularisation syndromes, scarring and uterine fibroids. They may
CC also be useful in wound healing and in contraception. The present
CC sequence represents a polypeptide of the invention.
XX
XX Sequence 300 AA;
SQ
Query Match 19.7%; Score 60; DB 7; Length 300;
Best Local Similarity 35.0%; Pred. No. 43;
Matches 14; Conservative 7; Mismatches 19; Indels 0; Gaps 0
QY 5 GTNGVHCTKNSLWGKVLKNNKYDQNTTYTMYGRLODILGW 44
DB 113 GTQVPCNKILLWSRIKDLAHQFTQVQRDMFTLEDTLGY 152
Search completed: May 16, 2005, 17:31:07
Job time : 166 secs

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2005, 17:27:38 ; Search time 22 Seconds
(without alignments)
183.230 Million cell updates/sec

Title: US-10-644-927-1
Perfect score: 304
Sequence: 1 KTYGTNGVHCTKNSLWGVK.....GRLQDILLGWATGAFGKTFH 54

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues
Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents AA:*

1: /cgn2_6/ptodata/1/iaa/5A COMB.pcp.*
2: /cgn2_6/ptodata/1/iaa/5B COMB.pcp.*
3: /cgn2_6/ptodata/1/iaa/6A COMB.pcp.*
4: /cgn2_6/ptodata/1/iaa/6B COMB.pcp.*
5: /cgn2_6/ptodata/1/iaa/PTUS COMB.pcp.*
6: /cgn2_6/ptodata/1/iaa/backfiles1.pcp.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69.5	22.9	420	4	US-09-328-352-5261
2	64	21.1	61	4	US-09-068-507C-5
3	62	20.4	85	4	US-09-270-767-44549
4	60	19.7	36	4	US-08-924-629C-69
5	60	19.7	37	4	US-09-030-619-209
6	60	19.7	61	4	US-08-924-629C-68
7	59.5	19.6	577	4	US-09-248-796A-18807
8	59	19.4	62	4	US-08-924-629C-71
9	56.5	18.6	242	4	US-09-543-681A-4928
10	56	18.4	66	4	US-09-107-532A-4680
11	56	18.4	10182	3	US-09-134-001C-3159
12	55.5	18.3	243	3	US-09-216-295-13
13	55.5	18.3	244	4	US-09-632-570-13
14	55.5	18.3	244	4	US-09-632-575-43
15	55	18.1	41	4	US-08-924-629C-70
16	55	18.1	177	4	US-09-248-796A-18521
17	54.5	17.9	494	3	US-08-378-313-23
18	54.5	17.9	494	3	US-08-378-313-29
19	54.5	17.9	641	4	US-09-071-035-456
20	54.5	17.9	1313	4	US-09-071-035-450
21	54.5	17.9	1313	4	US-09-071-035-454
22	54	17.8	697	4	US-09-489-039A-7485
23	53.5	17.6	56	4	US-09-834-309-3
24	53.5	17.6	74	4	US-09-107-532A-5289
25	53.5	17.6	134	4	US-09-834-309-4
26	53.5	17.6	1033	4	US-09-834-309-1
27	53	17.4	233	1	US-08-032-848C-12

28	53	17.4	263	3	US-09-216-295-24	Sequence 24, Appl
29	53	17.4	264	4	US-09-632-570-24	Sequence 24, Appl
30	53	17.4	264	4	US-09-632-575-54	Sequence 54, Appl
31	53	17.4	384	3	US-09-311-170-2	Sequence 2, Appl
32	53	17.4	929	4	US-09-134-000C-6424	Sequence 6424, A
33	52.5	17.3	371	4	US-09-248-796A-17748	Sequence 17748, A
34	52.5	17.3	418	4	US-09-610-104C-12	Sequence 12, Appl
35	52.5	17.3	911	3	US-09-356-952-6	Sequence 6, Appl
36	52.5	17.3	1176	1	US-08-257-999-2	Sequence 2, Appl
37	52	17.1	219	4	US-09-540-236-2633	Sequence 2633, Ap
38	52	17.1	221	4	US-09-270-767-33753	Sequence 33753, A
39	52	17.1	249	4	US-09-282-218A-17	Sequence 17, Appl
40	52	17.1	459	2	US-08-810-655A-2	Sequence 2, Appl
41	52	17.1	492	4	US-09-540-236-3061	Sequence 3061, Ap
42	51.5	16.9	139	4	US-09-902-540-13820	Sequence 13820, A
43	51.5	16.9	360	4	US-09-902-540-11642	Sequence 11642, A
44	51.5	16.9	510	4	US-09-727-628-2	Sequence 2, Appl
45	51	16.8	59	4	US-08-924-629C-74	Sequence 74, Appl

ALIGNMENTS

RESULT 1
US-09-328-352-5261
; Sequence 5261, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
; FILE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: GTC99-03PA
; CURRENT APPLICATION NUMBER: US/09/328,352
; CURRENT FILING DATE: 1999-06-04
; NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 5261
; LENGTH: 420
; TYPE: PRT
; ORGANISM: Acinetobacter baumannii
US-09-328-352-5261

Query Match 22.9%; Score 69.5; DB 4; Length 420;
Best Local Similarity 38.5%; Pred. No. 0.41;
Matches 20; Conservative 8; Mismatches 13; Indels 11; Gaps 4;
Qy 1 KTYGTN---GVHCTKNSLWGVRLKXMKY--DQNTYMGRLQDI---LLGW 44
Db 361 ESIVVTNPKSGKHSNKANIWIK---NNTPYSGDRDDTYTRRLRDITSPLIGW 409

RESULT 2
US-09-068-507C-5
; Sequence 5, Application US/09068507C
; Patent No. 6790951
; GENERAL INFORMATION:
; APPLICANT: ELSINK, VINCENT et al.
; TITLE OF INVENTION: EXPRESSION SYSTEM IN MICROORGANISM AND ITS USE FOR EXPRESSING
; FILE REFERENCE: 1380-0122P
; CURRENT APPLICATION NUMBER: US/09/068,507C
; CURRENT FILING DATE: 1998-07-15
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 5
; LENGTH: 61
; TYPE: PRT
; ORGANISM: Lactobacillus sake
US-09-068-507C-5

Query Match 21.1%; Score 64; DB 4; Length 61;
Best Local Similarity 36.7%; Pred. No. 0.2;
Matches 18; Conservative 4; Mismatches 9; Indels 18; Gaps 4;

Sequence 18807, Application US/09248796A
Patent No. 6747137
GENERAL INFORMATION:
APPLICANT: Keith Weinstock et al
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
FILE REFERENCE: 107196.132
CURRENT APPLICATION NUMBER: US/09/248,796A
PRIOR FILING DATE: 1999-02-12
PRIOR APPLICATION NUMBER: US 60/074,725
PRIOR FILING DATE: 1998-02-13
PRIOR APPLICATION NUMBER: US 60/096,409
PRIOR FILING DATE: 1998-08-13
NUMBER OF SEQ ID NOS: 28208
SEQ ID NO 18807
LENGTH: 577
TYPE: PRT
ORGANISM: Candida albicans
US-09-248-796A-18807

Query Match 19.6%; Score 59.5; DB 4; Length 577;
Best Local Similarity 30.6%; Pred. No. 14;
Matches 15; Conservative 7; Mismatches 12; Indels 15; Gaps 2;
QY 6 TNGVHCTKNSLWGVKRLKMKYDQNTTMYMGRLODILGWATGAFGKTFH 54
DB 261 TTGVQDDESTLWVKL-IPNLKHLNN-----QTSAGKRFH 294

RESULT 8
US-08-924-629C-71
Sequence 71, Application US/08924629C
Patent No. 6403082
GENERAL INFORMATION:
APPLICANT: Stiles, Michael E.
APPLICANT: Vederas, John C.
APPLICANT: van Belkum, Marius J.
APPLICANT: Worobo, Randy W.
APPLICANT: Worobo, Rodney J.
APPLICANT: Greer, G. Gordon
APPLICANT: McMullen, Lynn M.
APPLICANT: Leisner, Jorgen J.
APPLICANT: Poon, Alison
APPLICANT: Franz, Charles M.A.P.
TITLE OF INVENTION: No. 6403082elbacteriocins, Transport and Vector System and Method
FILE REFERENCE: 660.0005US
CURRENT APPLICATION NUMBER: US/08/924,629C
CURRENT FILING DATE: 1997-09-05
PRIOR APPLICATION NUMBER: US 60/026,257
PRIOR FILING DATE: 1996-09-05
NUMBER OF SEQ ID NOS: 80
SOFTWARE: Patentin version 3.1
SEQ ID NO 71
LENGTH: 62
TYPE: PRT
ORGANISM: Pediocin PA1
US-08-924-629C-71

Query Match 19.4%; Score 59; DB 4; Length 62;
Best Local Similarity 37.5%; Pred. No. 0.99;
Matches 18; Conservative 5; Mismatches 9; Indels 16; Gaps 4;
QY 3 YGTNGVHCTKNSL---WGVKRLKMKYDQNTTMYMGRLODILGWATG 47
DB 20 YYG-NGVTCGRKSCVDWGKA-----TTTCI---INNGAWATG 54

RESULT 9
US-09-543-681A-4928
Sequence 4928, Application US/09543681A
Patent No. 6605709
GENERAL INFORMATION:
APPLICANT: GARY BRETON

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
FILE REFERENCE: 2709.1002-001
CURRENT APPLICATION NUMBER: US/09/543,681A
CURRENT FILING DATE: 2000-04-05
PRIOR APPLICATION NUMBER: US 60/128,706
PRIOR FILING DATE: 1999-04-09
NUMBER OF SEQ ID NOS: 8344
SEQ ID NO 4928
LENGTH: 242
TYPE: PRT
ORGANISM: Proteus mirabilis
FEATURE:
NAME/KEY: UNSURE
LOCATION: (18)
OTHER INFORMATION: Identity of amino acid at the above locations are unknown.
US-09-543-681A-4928

Query Match 18.6%; Score 56.5; DB 4; Length 242;
Best Local Similarity 28.1%; Pred. No. 12;
Matches 18; Conservative 8; Mismatches 21; Indels 17; Gaps 3;
QY 4 YGTNGVHCTKNSLWGVKRLKN-----MKYDQNTTMYMGRLODILGWATGAF 49
DB 106 YKTSSTDSAKISVFLVNAKNTEQIGYIIDVIDIKVDNKRAYLGRVADKI--W-TGSS 162
QY 50 GKTF 53
DB 163 ATIF 166

RESULT 10
US-09-107-532A-4680
Sequence 4680, Application US/09107532A
Patent No. 6583275
GENERAL INFORMATION:
APPLICANT: Lynn A Doucette-Stamm and David Bush
TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
NUMBER OF SEQUENCES: 7310
CORRESPONDENCE ADDRESS:
ADDRESSEE: GENOME THERAPEUTICS CORPORATION
STREET: 100 Beaver Street
CITY: Waltham
STATE: Massachusetts
COUNTRY: USA
ZIP: 02354
COMPUTER READABLE FORM:
MEDIUM TYPE: CD-ROM ISO9660
COMPUTER: PC
OPERATING SYSTEM: <Unknown>
SOFTWARE: ASCII
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,532A
FILING DATE: 30-Jun-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/085,598
FILING DATE: 14 May 1998
APPLICATION NUMBER: 60/051571
FILING DATE: July 2, 1997
ATTORNEY/AGENT INFORMATION:
NAME: Ariniello, Pamela Deneke
REGISTRATION NUMBER: 40,489
REFERENCE/DOCKET NUMBER: GTC-012
TELECOMMUNICATION INFORMATION:
TELEPHONE: (781)893-5007
TELEFAX: (781)893-8277
INFORMATION FOR SEQ ID NO: 4680:
SEQUENCE CHARACTERISTICS:
LENGTH: 66 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

; APPLICANT: Worobo, Rodney J.
; APPLICANT: Greer, G. Gordon
; APPLICANT: McMullen, Lynn M.
; APPLICANT: Leisner, Jorgen J.
; APPLICANT: Poon, Aleson
; APPLICANT: Franz, Charles M.A.P.
; TITLE OF INVENTION: No. 6403082elBacteriocins, Transport and Vector System and Method
; FILE REFERENCE: 660.0005US
; CURRENT APPLICATION NUMBER: US/08/924,629C
; CURRENT FILING DATE: 1997-09-05
; PRIOR APPLICATION NUMBER: US 60/026,257
; PRIOR FILING DATE: 1996-09-05
; NUMBER OF SEQ ID NOS: 80
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 70
; LENGTH: 41
; TYPE: PRT
; ORGANISM: Sakacin P
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (33)..(33)
; OTHER INFORMATION: any amino acid
; NAME/KEY: MISC FEATURE
; LOCATION: (37)..(37)
; OTHER INFORMATION: any amino acid
; NAME/KEY: MISC FEATURE
; LOCATION: (40)..(40)
; OTHER INFORMATION: any amino acid
; NAME/KEY: MISC FEATURE
; LOCATION: (41)..(41)
; OTHER INFORMATION: any amino acid
US-08-924-629C-70

Query Match 18.1%; Score 55; DB 4; Length 41;
Best Local Similarity 63.2%; Pred. No. 2.1;
Matches 12; Conservative 1; Mismatches 2; Indels 4; Gaps 2;
QY 3 YGTGNGVHCTKNSL--WG 18
||| ||||| |:
Db 2 YYG-NGVHCGKHSCTVDWG 19

Search completed: May 16, 2005, 17:35:11
Job time : 23 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2005, 17:34:06 ; Search time 134 Seconds
(without alignments)
134.618 Million cell updates/sec

Title: US-10-644-927-1

Perfect score: 304

Sequence: 1 KTYGTNGVHCTKNSLWGKV.....GRLQDILLGWATGAFGKTFH 54

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1432185 seqs, 334051727 residues

*Total number of hits satisfying chosen parameters: 1432185

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

*Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications AA:*

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
- 2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep.*
- 3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
- 4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
- 5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
- 6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep.*
- 7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
- 8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
- 9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep.*
- 10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep.*
- 11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep.*
- 12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
- 13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep.*
- 14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep.*
- 15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep.*
- 16: /cgn2_6/ptodata/1/pubpaa/US10D_PUBCOMB.pep.*
- 17: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
- 18: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB.pep.*
- 19: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
- 20: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	60	19.7	36	10 US-09-883-343A-69	Sequence 69, Appl
2	60	19.7	37	9 US-09-030-619-209	Sequence 209, App
3	60	19.7	37	9 US-09-912-609-97	Sequence 97, Appl
4	60	19.7	37	15 US-10-277-232-209	Sequence 209, App
5	60	19.7	37	15 US-10-277-233-209	Sequence 209, App
6	60	19.7	61	10 US-09-883-343A-68	Sequence 68, Appl
7	60	19.7	300	9 US-09-982-616-9	Sequence 9, Appli
8	60	19.7	300	10 US-09-773-307B-2	Sequence 2, Appli
9	60	19.7	300	15 US-10-295-027-1348	Sequence 1348, Ap
10	60	19.7	300	15 US-10-454-238-2	Sequence 2, Appli
11	60	19.7	300	16 US-10-322-696-162	Sequence 162, App
12	60	19.7	300	17 US-10-753-267-34	Sequence 34, Appl
13	59	19.4	62	10 US-09-883-343A-71	Sequence 71, Appl

Sequence 6141, Ap
Sequence 280304,
Sequence 213009,
Sequence 10, Appl
Sequence 10, Appl
Sequence 197752,
Sequence 72532, A
Sequence 225665,
Sequence 11, Appl
Sequence 70580, A
Sequence 23, Appl
Sequence 13, Appl
Sequence 13, Appl
Sequence 70314, A
Sequence 60740, A
Sequence 70, Appl
Sequence 5582, Ap
Sequence 5591, A
Sequence 103806,
Sequence 456, App
Sequence 456, App
Sequence 450, App
Sequence 454, App
Sequence 450, App
Sequence 56879, A
Sequence 454, App
Sequence 454, App
Sequence 10936, A
Sequence 732, App
Sequence 3, Appli

ALIGNMENTS

RESULT 1

US-09-883-343A-69

; Sequence 69, Application US/09883343A

; Publication No. US20030039632A1

; GENERAL INFORMATION:

; APPLICANT: Stiles, Michael E.

; APPLICANT: Vederas, John C.

; APPLICANT: van Belkum, Marius J.

; APPLICANT: Worobo, Randy W.

; APPLICANT: Worobo, Rodney J.

; APPLICANT: Greer, G. Gordon

; APPLICANT: McMullen, Lynn M.

; APPLICANT: Leisner, Jorgen J.

; APPLICANT: Poon, Alston

; APPLICANT: Franz, Charles M.A.P.

; TITLE OF INVENTION: No. US20030039632A1eBacteriocins, Transport and Vector System and

; FILE REFERENCE: 660.0005US

; CURRENT APPLICATION NUMBER: US/09/883,343A

; CURRENT FILING DATE: 2001-06-19

; PRIOR APPLICATION NUMBER: US/08/924,629

; PRIOR FILING DATE: 1997-09-05

; PRIOR APPLICATION NUMBER: US 60/026,257

; PRIOR FILING DATE: 1996-09-05

; NUMBER OF SEQ ID NOS: 80

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 69

; LENGTH: 36

; TYPE: PRT

; ORGANISM: Mesenteriocin Y105

US-09-883-343A-69

Query Match 19.7%; Score 60; DB 10; Length 36;

Best Local Similarity 48.4%; Pred.No. 2.3;

Matches 15; Conservative 2; Mismatches 4; Indels 10; Gaps 3;

QY 3 YYGTNGVHCTKNSL---WGKV-----RLKN 24


```

; APPLICANT: BML, INC.
; TITLE OF INVENTION: Method of Detecting Risk Factor for Onset of Diabetes
; FILE REFERENCE: PBM37
; CURRENT APPLICATION NUMBER: US/09/773,307B
; CURRENT FILING DATE: 2001-01-31
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 2
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Hominidae
US-09-773-307B-2

Query Match      19.7%; Score 60; DB 10; Length 300;
Best Local Similarity 35.0%; Pred. No. 25;
Matches 14; Conservative 7; Mismatches 19; Indels 0; Gaps 0;

Qy      5 GTGNGVCTKNSLKGKVLKMKYKDQNTTYMGRLODILLGW 44
Db      113 GTQVPCNKILLWSRIKDLAHQFTQVQORDMFTLEDTLLG 152

RESULT 9
US-10-295-027-1348
; Sequence 1348, Application US/10295027
; Publication No. US20030232350A1
; GENERAL INFORMATION:
; APPLICANT: Afar, Daniel
; APPLICANT: Aziz, Natasha
; APPLICANT: Ginsberg, Wendy M.
; APPLICANT: Gish, Kurt C.
; APPLICANT: Glynn, Richard
; APPLICANT: Hevezi, Peter A.
; APPLICANT: Mack, David H.
; APPLICANT: Murray, Richard
; APPLICANT: Watson, Susan R.
; APPLICANT: Eos Biotechnology, Inc.
; TITLE OF INVENTION: Methods of Diagnosis of Cancer, Compositions and
; TITLE OF INVENTION: Methods of Screening for Modulators of Cancer
; FILE REFERENCE: 018501-012500US
; CURRENT APPLICATION NUMBER: US/10/295,027
; CURRENT FILING DATE: 2002-11-13
; PRIOR APPLICATION NUMBER: US 09/663,733
; PRIOR FILING DATE: 2000-09-15
; PRIOR APPLICATION NUMBER: US 60/350,666
; PRIOR FILING DATE: 2001-11-13
; PRIOR APPLICATION NUMBER: US 60/335,394
; PRIOR FILING DATE: 2001-11-15
; PRIOR APPLICATION NUMBER: US 60/332,464
; PRIOR FILING DATE: 2001-11-21
; PRIOR APPLICATION NUMBER: US 60/334,393
; PRIOR FILING DATE: 2001-11-29
; PRIOR APPLICATION NUMBER: US 60/340,376
; PRIOR FILING DATE: 2001-12-14
; PRIOR APPLICATION NUMBER: US 60/347,211
; PRIOR FILING DATE: 2002-01-08
; PRIOR APPLICATION NUMBER: US 60/347,349
; PRIOR FILING DATE: 2002-01-10
; PRIOR APPLICATION NUMBER: US 60/355,250
; PRIOR FILING DATE: 2002-02-08
; PRIOR APPLICATION NUMBER: US 60/356,714
; PRIOR FILING DATE: 2002-02-13
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 1386
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 1348
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-295-027-1348

Query Match      19.7%; Score 60; DB 15; Length 300;
Best Local Similarity 35.0%; Pred. No. 25;

```

```
Matches 14; Conservative 7; Mismatches 19; Indels 0; Gaps 0;
QY 5 GTNGVHCTKNSLWKGKVLKMKYDQNTYMGRIQDILLGW 44
Db 113 GTQTPCNKILLWSRIKDLAHQFTQVQORDMFTLEDTLGGY 152

RESULT 10
US-10-454-238-2
; Sequence 2, Application US/10454238
; Publication No. US20040081981A1
; GENERAL INFORMATION:
; APPLICANT: Eguashiro et al.
; TITLE OF INVENTION: Method of Detecting Risk Factor for Onset of Diabetes
; FILE REFERENCE: MSHING.001DV1
; CURRENT APPLICATION NUMBER: US/10/454,238
; PRIOR FILING DATE: 2003-06-04
; PRIOR FILING DATE: 2003-07-307
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Hominidae
US-10-454-238-2

Query Match 19.7%; Score 60; DB 15; Length 300;
Best Local Similarity 35.0%; Pred. No. 25;
Matches 14; Conservative 7; Mismatches 19; Indels 0; Gaps 0;
QY 5 GTNGVHCTKNSLWKGKVLKMKYDQNTYMGRIQDILLGW 44
Db 113 GTQTPCNKILLWSRIKDLAHQFTQVQORDMFTLEDTLGGY 152

RESULT 11
US-10-322-696-162
; Sequence 162, Application US/10322696
; Publication No. US20040166490A1
; GENERAL INFORMATION:
; APPLICANT: Morris, David W.
; APPLICANT: Malandro, Marc
; TITLE OF INVENTION: NOVEL THERAPEUTIC TARGETS IN CANCER
; FILE REFERENCE: 529452001200
; CURRENT APPLICATION NUMBER: US/10/322,696
; CURRENT FILING DATE: 2003-10-17
; NUMBER OF SEQ ID NOS: 186
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 162
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-322-696-162

Query Match 19.7%; Score 60; DB 16; Length 300;
Best Local Similarity 35.0%; Pred. No. 25;
Matches 14; Conservative 7; Mismatches 19; Indels 0; Gaps 0;
QY 5 GTNGVHCTKNSLWKGKVLKMKYDQNTYMGRIQDILLGW 44
Db 113 GTQTPCNKILLWSRIKDLAHQFTQVQORDMFTLEDTLGGY 152

RESULT 12
US-10-753-267-34
; Sequence 34, Application US/10753267
; Publication No. US20050037946A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Stagliano, Nancy E.
; APPLICANT: Healy, Aileen
; APPLICANT: Acton, Susan L.
```

```
; APPLICANT: Galvin, Katherine M.
; APPLICANT: Donoghue, Mary A.
; APPLICANT: Rodriquez-Way, Amelie
; APPLICANT: Tomlinson, James E.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING
; TITLE OF INVENTION: CARDIOVASCULAR DISEASE USING 1722, 10280, 59917, 85553,
; TITLE OF INVENTION: 10653, 9235, 21668, 17794, 2210, 6169, 10102, 21061, 17662,
; TITLE OF INVENTION: 1468, 12282, 6350, 9035, 1820, 23652, 7301, 8925, 8701,
; TITLE OF INVENTION: 3533, 9462, 9123, 12788, 17729, 65552, 1261, 21476, 33770,
; TITLE OF INVENTION: 9380, 2569654, 33556, 53656, 44143, 32612, 10671, 261,
; TITLE OF INVENTION: 44570, 41922, 2552, 2417, 19319, 43969, 8921, 8993, 955,
; TITLE OF INVENTION: 32345, 966, 1920, 17318, 1510, 14180, 26005, 554, 16408,
; TITLE OF INVENTION: 42028, 112091, 13886, 13942, 1673, 54946 OR 2419
; FILE REFERENCE: MPI03-003PIRNOVMIM
; CURRENT APPLICATION NUMBER: US/10/753,267
; CURRENT FILING DATE: 2004-01-08
; PRIOR APPLICATION NUMBER: US 60/439,683
; PRIOR FILING DATE: 2003-01-13
; PRIOR APPLICATION NUMBER: US 60/445,216
; PRIOR FILING DATE: 2003-02-05
; PRIOR APPLICATION NUMBER: US 60/448,036
; PRIOR FILING DATE: 2003-02-18
; PRIOR APPLICATION NUMBER: US 60/454,189
; PRIOR FILING DATE: 2003-03-12
; PRIOR APPLICATION NUMBER: US 60/457,541
; PRIOR FILING DATE: 2003-03-25
; PRIOR APPLICATION NUMBER: US 60/466,411
; PRIOR FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: US 60/469,041
; PRIOR FILING DATE: 2003-05-08
; PRIOR APPLICATION NUMBER: US 60/477,414
; PRIOR FILING DATE: 2003-06-10
; PRIOR APPLICATION NUMBER: US 60/478,560
; PRIOR FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/489,772
; PRIOR FILING DATE: 2003-07-24
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 130
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 300
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-753-267-34

Query Match 19.7%; Score 60; DB 17; Length 300;
Best Local Similarity 35.0%; Pred. No. 25;
Matches 14; Conservative 7; Mismatches 19; Indels 0; Gaps 0;
QY 5 GTNGVHCTKNSLWKGKVLKMKYDQNTYMGRIQDILLGW 44
Db 113 GTQTPCNKILLWSRIKDLAHQFTQVQORDMFTLEDTLGGY 152

RESULT 13
US-09-883-343A-71
; Sequence 71, Application US/09883343A
; Publication No. US20030039632A1
; GENERAL INFORMATION:
; APPLICANT: Stiles, Michael E.
; APPLICANT: Vederas, John C.
; APPLICANT: van Belkum, Marius J.
; APPLICANT: Worobo, Rodney W.
; APPLICANT: Greer, G. Gordon
; APPLICANT: McMullen, Lynn M.
; APPLICANT: Leisner, Joergen J.
; APPLICANT: Poon, Alston
; APPLICANT: Franz, Charles M.A.P.
; TITLE OF INVENTION: No. US20030039632A1e1Bacteriocins, Transport and Vector System an
; FILE REFERENCE: 660.0005US
; CURRENT APPLICATION NUMBER: US/09/883,343A
; CURRENT FILING DATE: 2001-06-19
```

```

RESULT 15
US-10-424-599-280304
; Sequence 280304, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 280304
; LENGTH: 95
; TYPE: PRT
; ORGANISM: Glycine max

```

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2005, 17:25:23 ; Search time 39 Seconds
(without alignments)
133.223 Million cell updates/sec

Title: US-10-644-927-1
Perfect score: 304
Sequence: 1 KTYVGTNGVHCTKNSLWGKV.....GRLQDILLGWATGAFGKTFH 54
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 79:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	65.5	21.5	471	2 S11899	glutamate-ammonia
2	65.5	21.5	471	2 A99180	hypothetical prote
3	64	21.1	61	1 S38508	sakacin P precursor
4	63	20.7	590	2 AF2555	hypothetical prote
5	62.5	20.6	446	2 A43995	glutamate-ammonia
6	60	19.7	61	2 A41657	leucocin A-UAL 187
7	60	19.7	61	2 S52208	mesentericin Y105
8	60	19.7	300	2 A43521	lymphocyte surface
9	59.5	19.6	286	2 F84554	spermidine synthas
10	59	19.4	62	1 A48941	pediocin PA-1 prec
11	59	19.4	399	2 T20455	hypothetical prote
12	58	19.1	3036	2 T18995	hypothetical prote
13	57.5	18.9	510	2 T50921	inositol-3-phospha
14	56.5	18.6	517	1 G1BPT4	gene 12 protein -
15	55.5	18.3	132	2 FC2131	hepatocyte growth
16	55.5	18.3	934	2 B29838	parasporeal crystal
17	55.5	18.3	1332	2 F69732	PBSX prophage ORF
18	55	18.1	201	2 AD3302	31K outer-membrane
19	55	18.1	346	2 B90243	aminomethyltransfe
20	55	18.1	369	2 T24205	hypothetical prote
21	55	18.1	516	2 G84442	probable nucleosid
22	55	18.1	799	1 TVRTT8	nerve growth facto
23	54.5	17.9	319	2 T15463	hypothetical prote
24	54.5	17.9	322	2 D86760	phosphate starvati
25	54.5	17.9	374	2 F97257	glycosyltransferas
26	54.5	17.9	383	2 B96806	hypothetical prote
27	54.5	17.9	494	1 B41141	1-aminocyclopropan
28	54	17.8	262	1 TLBPM1	tail fiber protein
29	54	17.8	269	2 S51815	hypothetical prote

30	54	17.8	494	2 JC2382	sodium/proline sym
31	53.5	17.6	254	2 S31393	chlorophyll a/b-bi
32	53.5	17.6	324	2 G69515	transcription regu
33	53.5	17.6	462	1 YSBVTM	threonine-tRNA lig
34	53.5	17.6	510	2 D84610	probable myo-inosi
35	53.5	17.6	527	2 S29530	gene 12 protein -
36	53.5	17.6	1091	1 PL0009	complement C3d/Bps
37	53.5	17.6	1166	2 H71609	hypothetical prote
38	53.5	17.6	1585	2 B69948	phage-related prot
39	53	17.4	167	2 S49628	FN81 protein - ye
40	53	17.4	202	2 C90228	hypothetical prote
41	53	17.4	249	2 B95357	probable EtfB2 ele
42	53	17.4	258	2 T13591	tail fiber adhesin
43	53	17.4	264	2 JU0328	cellulase (EC 3.2.
44	53	17.4	327	2 AI3308	glutathione transf
45	53	17.4	327	2 S54560	TOM37 protein - ye

ALIGNMENTS

RESULT 1

S11899

Glutamate-ammonia ligase (EC 6.3.1.2) - Sulfolobus solfataricus

N;Alternate names: Glutamine synthetase

C;Species: Sulfolobus solfataricus

C;Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 09-Jul-2004

C;Accession: S11899

R;Sanangelanton, A.M.; Barbarini, D.; di Pasquale, G.; Cammarano, P.; Tiboni, O.

Mol. Gen. Genet. 221, 187-194, 1990

A;Title: Cloning and nucleotide sequence of an archaeobacterial glutamine synthetase gene

A;Reference number: S11899; MUID:90318316; PMID:1973523

A;Accession: S11899

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-471 <SAN>

A;Cross-references: UNIPROT:P23794; GB:X53263; NID:g296864; PIDN:CAA37353.1; PID:g296865

C;Superfamily: Glutamate-ammonia ligase

C;Keywords: ligase

Query Match 21.5%; Score 65.5; DB 2; Length 471;
Best Local Similarity 48.6%; Pred. No. 1.8;
Matches 17; Conservative 2; Mismatches 13; Indels 3; Gaps 2;

QY 1 KTYVGTN--GVHCTKNSLWGKVLKNMKYDONTTY 33
DB 255 KPIYDNGTGMH-THLSLWTKDGKKNLMYDPNDEY 288

RESULT 2

A99180

hypothetical protein glnA-1 [imported] - Sulfolobus solfataricus

C;Species: Sulfolobus solfataricus

C;Date: 24-May-2001 #sequence_revision 24-May-2001 #text_change 09-Jul-2004

C;Accession: A99180

R;She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Awayez, M.J.; Chan;

Jong, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, P.

arrett, R.A.; Ragan, M.A.; Sengen, C.W.; Van der Oost, J.

submitted to GenBank, April 2001

A;Description: Sulfolobus solfataricus complete genome.

A;Reference number: A99139

A;Accession: A99180

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-471 <KOR>

A;Cross-references: UNIPROT:P23794; GB:AB006641; NID:g13813510; PIDN:AAK40696.1; GSPDB:G

C;Genetics:

A;Gene: glnA-1

C;Superfamily: glutamate-ammonia ligase

Query Match 21.5%; Score 65.5; DB 2; Length 471;
Best Local Similarity 48.6%; Pred. No. 1.8;
Matches 17; Conservative 2; Mismatches 13; Indels 3; Gaps 2;

A;Residues: 1-590 <KUR>
A;Cross-references: UNIPROT:Q8YK80; GB:AP003603; PIDN:BAB77367.1; PID:g17134810; GSPDB:G17134810
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all18037
A;Genome: plasmid

Query Match 20.7%; Score 63; DB 2; Length 590;
Best Local Similarity 35.3%; Pred. No. 5;
Matches 18; Conservative 8; Mismatches 23; Indels 2; Gaps 1;

QY 3 YGTGNGVHCTKNSLWGKVRLEKKNKYDQNTTYMGRLODILLGWATGAFGKTF 53
DB 88 YGTG--PRGTFQVGVGNKRNITNIPEDKNRLYLDPDVGILVSGSGSGKTF 136

RESULT 3
A43995
glutamate-ammonia ligase (EC 6.3.1.2) - Methanococcus voltae
N;Alternate names: glutamine synthetase
C;Species: Methanococcus voltae
C;Date: 11-Feb-1993 #sequence_revision 11-Feb-1993 #text_change 09-Jul-2004
C;Accession: A43995; S13504
R;Posot, O.; Sibold, L.; Aubert, J.P.
Res. Microbiol. 140, 355-371, 1989
A;Title: Nucleotide sequence and expression of the glutamine synthetase structural gene,
A;Reference number: A43995; MUID:90139872; PMID:2575777
A;Accession: A43995
A;Molecule type: DNA
A;Residues: 1-446 <POS>
A;Cross-references: UNIPROT:P21154; GB:X53509; NID:g44711; PIDN:CAA37585.1; PID:g44712
C;Genetics:
A;Gene: glnA
C;Superfamily: glutamate-ammonia ligase
C;Keywords: ligase

Query Match 20.6%; Score 62.5; DB 2; Length 446;
Best Local Similarity 33.9%; Pred. No. 4.2;
Matches 19; Conservative 9; Mismatches 17; Indels 11; Gaps 5;

QY 1 KTYVGTN--GVHCTKNSLW--GKVRLEKKNKYDQNTTYMGRLODILLGWATGAFGKT 52
DB 237 KPFFGNGSGMHGNCQ-SIWLDKGKPSF----YDENNAH--QLSDICLSYIGILEHT 285

RESULT 6
A41657
leucocin A-UAL 187 precursor - Leuconostoc gelidium plasmid pLG7.6
C;Species: Leuconostoc gelidium
C;Date: 13-May-1992 #sequence_revision 13-May-1992 #text_change 09-Jul-2004
C;Accession: A41657
R;Hastings, J.W.; Sailer, M.; Johnson, K.; Roy, K.L.; Vederas, J.C.; Stiles, M.E.
J. Bacteriol. 173, 7491-7500, 1991
A;Title: Characterization of leucocin A-UAL 187 and cloning of the bacteriocin gene from
A;Reference number: A41657; MUID:92041660; PMID:1840587
A;Accession: A41657
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-61 <HAS>
A;Cross-references: UNIPROT:P34034; GB:M64371; NID:g149637; PIDN:AAA68003.1; PID:g149638
C;Genetics:
A;Genome: plasmid
C;Superfamily: carnobacteriocin precursor
C;Keywords: bacteriocin

Query Match 19.7%; Score 60; DB 2; Length 61;
Best Local Similarity 48.4%; Pred. No. 0.92;
Matches 15; Conservative 2; Mismatches 4; Indels 10; Gaps 3;

QY 3 YGTGNGVHCTKNSL---WGKVRLEKKNKYDQNTTYMGRLODILLGWATGAFGKT 24
DB 26 YYG-NGVHCTKSGCSVNWGEAFSAGVHRLAN 55

QY 1 KTYVGTN--GVHCTKNSLWGKVRLEKKNKYDQNTTY 33
DB 255 KPIYGDNGTGWH--THLSLWTKDGKKNLMLYDPNDEY 288

RESULT 3
S38508
sakacin P precursor - Lactobacillus sake (strains Lb674 and LTH673)
C;Species: Lactobacillus sake
A;Variety: strains Lb 674 and LTH 673
C;Date: 25-Dec-1994 #sequence_revision 13-Mar-1997 #text_change 09-Jul-2004
C;Accession: S57911; S57915; S43689; S38508
R;Huehne, K.; Holck, A.; Axelsson, L.; Kroeckel, L.
submitted to the EMBL Data Library, February 1995
A;Description: Purification and cloning of sakacin 674, a bacteriocin from Lactobacillus
A;Reference number: S57910
A;Accession: S57911
A;Molecule type: DNA
A;Residues: 1-61 <HUE>
A;Cross-references: UNIPROT:P35618; EMBL:Z48542; NID:g695615; PIDN:CAA84828.1; PID:g6956
A;Experimental source: strain Lb674
R;Holck, A.L.; Axelsson, L.; Huehne, K.; Kroeckel, L.
FEMS Microbiol. Lett. 115, 143-150, 1994
A;Title: Purification and cloning of sakacin 674, a bacteriocin from Lactobacillus sake
A;Reference number: S57915; MUID:94186010; PMID:8138128
A;Accession: S57915
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-61 <HOI>
A;Cross-references: EMBL:Z25816; NID:g414124; PIDN:CAA81064.1; PID:g414125
A;Experimental source: strain Lb674
R;Tichacek, P.S.; Vogel, R.F.; Hammes, W.P.
Microbiology 140, 361-367, 1994
A;Title: Cloning and sequencing of sakP encoding sakacin P, the bacteriocin produced by
A;Reference number: S43688; MUID:94236240; PMID:8180701
A;Accession: S43689
A;Molecule type: DNA
A;Residues: 1-61 <TIC>
A;Cross-references: EMBL:X75081; NID:g475950; PIDN:CAA52974.1; PID:g475952
A;Experimental source: strain LTH 673
C;Genetics:
A;Gene: sakP; sppA
C;Superfamily: carnobacteriocin precursor
C;Keywords: antibacterial; bacteriocin
F;1-18/Domain: propeptide #status predicted <PRO>
F;19-61/Product: sakacin P #status predicted <MAT>
F;27-32/Diulfide bonds: #status predicted

Query Match 21.1%; Score 64; DB 1; Length 61;
Best Local Similarity 36.7%; Pred. No. 0.28;
Matches 18; Conservative 4; Mismatches 9; Indels 18; Gaps 4;

QY 3 YGTGNGVHCTKNSL---WGKVRLEKKNKYDQNTTYMGRLODILLGWATG 47
DB 20 YYG-NGVHCTKSGKHSCTVDWG-----TAIGNIGNNAANWATG 54

RESULT 4
AF2555
hypothetical protein all18037 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120ga
C;Species: Nostoc sp. PCC 7120
A;Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C;Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 09-Jul-2004
C;Accession: AF2555
R;Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriiguchi,
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A;Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana
A;Reference number: AB1807; MUID:21595285; PMID:11759840
A;Accession: AF2555
A;Status: preliminary
A;Molecule type: DNA

A;Residues: 1-590 <KUR>
A;Cross-references: UNIPROT:Q8YK80; GB:AP003603; PIDN:BAB77367.1; PID:g17134810; GSPDB:G17134810
A;Experimental source: strain PCC 7120
C;Genetics:
A;Gene: all18037
A;Genome: plasmid

Query Match 20.7%; Score 63; DB 2; Length 590;
Best Local Similarity 35.3%; Pred. No. 5;
Matches 18; Conservative 8; Mismatches 23; Indels 2; Gaps 1;

QY 3 YGTGNGVHCTKNSLWGKVRLEKKNKYDQNTTYMGRLODILLGWATGAFGKTF 53
DB 88 YGTG--PRGTFQVGVGNKRNITNIPEDKNRLYLDPDVGILVSGSGSGKTF 136

RESULT 5
A43995
glutamate-ammonia ligase (EC 6.3.1.2) - Methanococcus voltae
N;Alternate names: glutamine synthetase
C;Species: Methanococcus voltae
C;Date: 11-Feb-1993 #sequence_revision 11-Feb-1993 #text_change 09-Jul-2004
C;Accession: A43995; S13504
R;Posot, O.; Sibold, L.; Aubert, J.P.
Res. Microbiol. 140, 355-371, 1989
A;Title: Nucleotide sequence and expression of the glutamine synthetase structural gene,
A;Reference number: A43995; MUID:90139872; PMID:2575777
A;Accession: A43995
A;Molecule type: DNA
A;Residues: 1-446 <POS>
A;Cross-references: UNIPROT:P21154; GB:X53509; NID:g44711; PIDN:CAA37585.1; PID:g44712
C;Genetics:
A;Gene: glnA
C;Superfamily: glutamate-ammonia ligase
C;Keywords: ligase

Query Match 20.6%; Score 62.5; DB 2; Length 446;
Best Local Similarity 33.9%; Pred. No. 4.2;
Matches 19; Conservative 9; Mismatches 17; Indels 11; Gaps 5;

QY 1 KTYVGTN--GVHCTKNSLW--GKVRLEKKNKYDQNTTYMGRLODILLGWATGAFGKT 52
DB 237 KPFFGNGSGMHGNCQ-SIWLDKGKPSF----YDENNAH--QLSDICLSYIGILEHT 285

RESULT 6
A41657
leucocin A-UAL 187 precursor - Leuconostoc gelidium plasmid pLG7.6
C;Species: Leuconostoc gelidium
C;Date: 13-May-1992 #sequence_revision 13-May-1992 #text_change 09-Jul-2004
C;Accession: A41657
R;Hastings, J.W.; Sailer, M.; Johnson, K.; Roy, K.L.; Vederas, J.C.; Stiles, M.E.
J. Bacteriol. 173, 7491-7500, 1991
A;Title: Characterization of leucocin A-UAL 187 and cloning of the bacteriocin gene from
A;Reference number: A41657; MUID:92041660; PMID:1840587
A;Accession: A41657
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-61 <HAS>
A;Cross-references: UNIPROT:P34034; GB:M64371; NID:g149637; PIDN:AAA68003.1; PID:g149638
C;Genetics:
A;Genome: plasmid
C;Superfamily: carnobacteriocin precursor
C;Keywords: bacteriocin

Query Match 19.7%; Score 60; DB 2; Length 61;
Best Local Similarity 48.4%; Pred. No. 0.92;
Matches 15; Conservative 2; Mismatches 4; Indels 10; Gaps 3;

QY 3 YGTGNGVHCTKNSL---WGKVRLEKKNKYDQNTTYMGRLODILLGWATGAFGKT 24
DB 26 YYG-NGVHCTKSGCSVNWGEAFSAGVHRLAN 55

RESULT 9
F84954
spermidine synthase [EC 2.5.1.16] [imported] - Buchnera sp. (strain APS)
C/Species: Buchnera sp.
C/Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C/Accession: F84954
R/Shigenobu, S.; Watanabe, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.
Nature 407, 81-86, 2000
A/Title: Genome sequence of the endocellular bacterial symbiont of aphids Buchnera sp. A
A/Reference number: A84930; UID:20445173; PMID:10993077
A/Accession: F84954
A>Status: preliminary
A/Molecule type: DNA
A/Residues: 1-286 <STO>
A/Cross-references: GB:AP000398; GSPDB:GN00144
A/Experimental source: strain APS
C/Genetics:
A/Gene: speE; BU209
C/Superfamily: spermidine synthase
C/Keywords: transference

Query Match 19.6%; Score 59.5; DB 2; Length 286;
Best Local Similarity 42.1%; Pred. No. 6.2;
Matches 16; Conservative 4; Mismatches 9; Indels 9; Gaps 2;

QY 2 TTY-----GTGNGVHCTKNSLWG-KVRLKNMKYDQN 30
||| : ||| : ||| : ||| :
DB 225 TTYGGVMVFAMGTNIEYRKNSLEKIQRINKTKLDNF 262

RESULT 10
A48941
pediocin PA-1 precursor - Pediococcus acidilactici plasmid pSRQ11
C/Species: Pediococcus acidilactici
C/Date: 19-Dec-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C/Accession: A48941; S21699; A47680; A41018; A41823
R/Marugg, J.D.; Gonzalez, C.F.; Kunka, B.S.; Ledesboer, A.M.; Pucci, M.J.; Toonen, M.Y.; V
Appl. Environ. Microbiol. 58, 2360-2367, 1992
A/Title: Cloning, expression, and nucleotide sequence of genes involved in production of
A/Reference number: A48941; UID:92384551; PMID:1514784
A/Accession: A48941
A/Molecule type: DNA
A/Residues: 1-62 <MAP>
A/Cross-references: UNIPROT:P28430; GB:M83924; NID:gl50565; PIDN:AAA25559.1; PID:gl50566
A/Experimental source: strain PAC1.0, plasmid pSRQ11
A/Note: sequence extracted from NCBI backbone (NCBIN:112082, NCBIP:112083)
R/Henderson, J.T.; Chopko, A.L.; van Wassenaer, P.D.
Arch. Biochem. Biophys. 295, 5-12, 1992
A/Title: Purification and primary structure of pediocin PA-1 produced by *Pediococcus aci*
A/Reference number: S21699; UID:92246549; PMID:1573516
A/Accession: S21699
A/Molecule type: protein
A/Residues: 19-62 <HEN>
R/Nieto Lozano, J.C.; Meyer, J.N.; Sletten, K.; Pelaz, C.; Nes, I.F.
J. Gen. Microbiol. 138, 1985-1990, 1992
A/Title: Purification and amino acid sequence of a bacteriocin produced by *Pediococcus a*
A/Reference number: A47680; UID:93019000; PMID:1402795
A/Accession: A47680
A/Molecule type: protein
A/Residues: 19-50, 'X', 52-61, 'X' <NIE>
A/Note: sequence extracted from NCBI backbone (NCBIP:116654)
A/Note: the unidentified residues were suggested to be Trp and Cys
C/Genetics:
A/Gene: peda
A/Genome: plaamid
C/Superfamily: carnobacteriocin precursor
C/Keywords: antibacterial; bacteriocin
F:1-16/Domain: propeptide #status predicted <PRO>
F:19-62/Product: pediocin PA-1 #status experimental <MAT>
F:37-32, 42-62/Disulfide bonds: #status experimental

```
Query Match      19.4%; Score 59; DB 1; Length 62;
Best Local Similarity 37.5%; Pred. No. 1.3;
Matches 18; Conservative 5; Mismatches 9; Indels 16; Gaps 4;

QY 3 YGTNGVHCTKNSL---WGKVRLLKNMKYDQNTTYMGRLODILLGWATG 47
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    : : : : : : : : : : : : : : : : : : : : : : : :
Db 20 YYG-NGVTCGKHSCSDVMGKA-----TTCTI--INNGMAAWATG 54

RESULT 11
T20455
hypothetical protein F01D4.3 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 12-Jul-2004
C:Accession: T20455
R:Wild, A.
submitted to the EMBL Data Library, October 1996
A:Reference number: Z19278
A:Accession: T20455
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-399 <WIL>
A:Cross-references: UNIPROT:O17755; EMBL:Z81054; PIDN: CAB02882.1; GSPDB:GN00022; CESP:FO
A:Experimental source: clone F01D4
C:Genetics:
A:Gene: CESP:F01D4.3
A:Map position: 4
A:Introns: 45/3; 235/2; 294/2

Query Match      19.4%; Score 59; DB 2; Length 399;
Best Local Similarity 35.8%; Pred. No. 10;
Matches 19; Conservative 4; Mismatches 22; Indels 8; Gaps 3;

QY 3 YGTNGVHCTKNSLWGVKVRLLKNMKYDQNTTYMGRLODILLG--WATGAFGKTF 53
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    : : : : : : : : : : : : : : : : : : : : : : : :
Db 75 YTTDNGSGCSN---GRFQLKNPLQQPWEYMH--SDVTAGKVLGEGAFGKVF 121

RESULT 12
T18995
hypothetical protein C06B8.7 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T18995
R:Steward, C.
submitted to the EMBL Data Library, November 1996
A:Reference number: Z19057
A:Accession: T18995
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3036 <WIL>
A:Cross-references: UNIPROT:O17575; EMBL:Z81463; PIDN: CAB03852.1; GSPDB:GN00023; CESP:CO
A:Experimental source: clone C06B8
C:Genetics:
A:Gene: CESP:C06B8.7
A:Map position: 5
A:Introns: 102/3; 157/2; 288/2; 439/3; 542/1; 591/3; 798/3; 835/3; 1139/1; 1204/3; 1325/

Query Match      19.1%; Score 58; DB 2; Length 3036;
Best Local Similarity 26.1%; Pred. No. 1.4e+02;
Matches 18; Conservative 5; Mismatches 24; Indels 22; Gaps 2;

QY 1 KTYGTNGVHCT-----KNSLWGVKVRLLKNMKYDQNTTYMGRLL-----Q 38
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    : : : : : : : : : : : : : : : : : : : : : : : :
Db 543 KAYVGLNVNTCIETKNTGNGVFANDIRERTALTNTVLTENQGVAGFLVKDGAADIWLNET 602

QY 39 DILLGWATG 47
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    : : : : : : : : : : : : : : : : : : : : : : : :
Db 603 RILNNMGDG 611

RESULT 13
T50021
inositol-3-phosphate synthase (EC 5.5.1.4) T31P16.160 [similarity] - Arabidopsis thaliana
N:Alternate names: protein T31P16.160
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C:Accession: T50021
R:Bevan, M.; Zimmermann, W.; Grueniseen, A.; Wambutt, R.; Kalicki, J.; Wohlmann, P.; Smj
submitted to the Protein Sequence Database, May 2000
A:Reference number: Z25027
A:Accession: T50021
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-510 <BEV>
A:Cross-references: UNIPROT:Q5LX12; EMBL:AL356332; GSPDB:GN00063; ATSP:T31P16.160
A:Experimental source: cultivar Columbia; BAC clone T31P16
C:Genetics:
A:Gene: ATSP:T31P16.160
A:Map position: 5
A:Introns: 63/2; 86/2; 131/3; 214/2; 290/1; 328/3; 387/3; 450/3; 471/3
C:Superfamily: myo-inositol-1-phosphate synthase
C:Keywords: intramolecular lyase; isomerase; NAD

Query Match      18.9%; Score 57.5; DB 2; Length 510;
Best Local Similarity 37.7%; Pred. No. 22;
Matches 20; Conservative 8; Mismatches 20; Indels 5; Gaps 4;

QY 3 YGTNGVHCTKNSL--WGKVRLLKNMKYDQNT--TYMGRLODILLGWATGAFGKT 52
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    : : : : : : : : : : : : : : : : : : : : : : : :
Db 26 YQTTELHVENKNGAFQM--TVKPKTVKYEFTDTHVPKLVGMVLYGVWG--GNNGST 76

RESULT 14
GIBFT4
gene 12 protein - phage T4
N:Alternate names: tail fiber protein
C:Species: phage T4
C:Date: 30-Sep-1991 #sequence_revision 30-Sep-1991 #text_change 30-Jun-1993
C:Accession: S01889
R:Selivanov, N.A.; Prilipov, A.G.; Mesyanzhinov, V.V.
Nucleic Acids Res. 16, 2334, 1988
A:Title: Nucleotide and deduced amino acid sequence of bacteriophage T4 gene 12.
A:Reference number: S01889; MUID:88189824; PMID:3357780
A:Accession: S01889
A:Molecule type: DNA
A:Residues: 1-517 <SEL>
A:Cross-references: EMBL:X06792
A:Note: the authors translated the codon CAG for residue 279 as His
C:Genetics:
A:Gene: 12
C:Superfamily: phage T4 gene 12 protein
C:Keywords: tail fiber

Query Match      18.6%; Score 56.5; DB 1; Length 517;
Best Local Similarity 33.3%; Pred. No. 30;
Matches 15; Conservative 8; Mismatches 15; Indels 7; Gaps 2;

QY 8 GVHCTKNSLWGVKVRLLKNMKYDQNTTYMGRLODILLGWATGAFGKT 52
    ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
    : : : : : : : : : : : : : : : : : : : : : : : :
Db 416 GVGCTGGYV-GEVQIQMSYHKHAGGFGHEDDL-----GAFGNT 453

RESULT 15
PC2131
hepatocyte growth factor receptor - rat (fragment)
C:Species: Rattus norvegicus (Norway rat)
C:Date: 03-May-1994 #sequence_revision 07-Oct-1994 #text_change 09-Jul-2004
C:Accession: PC2131
R:Tsuji, M.; Kawano, S.; Tsuji, S.; Ito, T.; Hayashi, N.; Horimoto, M.; Mita, E.; Nagas
Biochem. Biophys. Res. Commun. 200, 536-541, 1994
A:Title: Increased expression of c-met messenger RNA following acute gastric injury in r
A:Reference number: PC2131; MUID:94220137; PMID:8166728
A:Accession: PC2131
A:Molecule type: mRNA
A:Residues: 1-132 <TSU>
```

Search completed: May 16, 2005, 17:34:43
Job time : 41 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 16, 2005, 17:17:02 ; Search time 167 Seconds
(without alignments)
165.582 Million cell updates/sec

Title: US-10-644-927-1

Perfect score: 304

Sequence: 1 KTYGTNGVHCTKNSLWGVK.....GRLQDILLGWATGAFGKTFH 54

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

UniProt_03:*

1: uniprot_sprot:*

2: uniprot_trembl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	195	64.1	81	2 Q48496	Q48496 lactobacilli
2	65.5	21.5	471	1 GLNA SULSO	P23794 sulfolobus
3	64.5	21.2	1071	2 Q75C66	Q75C66 ashbya goss
4	64	21.1	61	1 SAKP_LACSK	P35618 lactobacilli
5	63	20.7	590	2 Q8YK80	Q8YK80 anabaena sp
6	63	20.7	1582	2 Q8RH71	Q8RH71 fusobacteri
7	63	20.7	1630	2 Q8RHH7	P21154 methanococ
8	62.5	20.6	446	1 GLNA METVO	Q8VW63 bacillus th
9	62.5	20.6	558	2 Q8VW63	Q8VW63 bacillus th
10	61.5	20.2	469	2 Q9P203	Q9P203 xestia c-ni
11	61.5	20.2	547	2 Q8A1J6	Q8A1J6 bacteroides
12	61.5	20.2	573	2 Q7QBD6	Q7QBD6 anopheles g
13	61.5	20.2	853	2 Q9DL93	Q9DL93 human immun
14	61.5	20.2	858	2 Q9DL79	Q9DL79 human immun
15	60	19.7	61	1 LCCA LEUGE	P34034 leuconostoc
16	60	19.7	61	1 LCCB LEUCA	Q53446 leuconostoc
17	60	19.7	61	1 MTCY LEUME	P38577 leuconostoc
18	60	19.7	61	2 Q79IV9	Q79IV9 leuconostoc
19	60	19.7	300	1 CD38 HUMAN	P28907 homo sapien
20	60	19.7	300	2 Q96HY4	Q96HY4 homo sapien
21	60	19.7	1143	2 Q948Y6	Q948Y6 volvox cart
22	60	19.7	1794	2 Q8RHH1	Q8RHH1 fusobacteri
23	60	19.7	1868	2 Q8IAL9	Q8IAL9 plasmodium
24	59.5	19.6	42	1 BAVM LACSK	P80493 lactobacilli
25	59.5	19.6	286	1 SPEE BUCAI	P57305 buchneza ap
26	59	19.4	62	1 PPAL_PEDAC	P29430 pediococcus
27	59	19.4	62	2 Q8RL96	Q8RL96 pediococcus
28	59	19.4	62	2 Q53278	Q53278 pediococcus
29	59	19.4	62	2 Q79IK0	Q79IK0 lactobacilli
30	59	19.4	62	2 Q79IK3	Q79IK3 pediococcus
31	59	19.4	62	2 Q79IK6	Q79IK6 pediococcus

32 59 19.4 62 2 Q9EZB2 Q9ezb2 bacillus co
33 59 19.4 67 2 Q47778 Q47778 enterococcu
34 59 19.4 299 2 Q65IN6 Q65in6 bacillus li
35 59 19.4 419 2 Q17755 Q17755 caenorhabdi
36 59 19.4 498 2 Q9N502 Q9n502 caenorhabdi
37 59 19.4 860 2 Q8WTM9 Q8wtm9 caenorhabdi
38 59 19.4 2589 2 Q7YYI7 Q7yyi7 cryptospori
39 58.5 19.2 66 2 Q9Z4J1 Q9z4j1 carnobacter
40 58.5 19.2 485 2 Q94I16 Q94i16 cucurbita m
41 58.5 19.2 853 2 Q9DL89 Q9dl89 human immun
42 58.5 19.2 853 2 Q9DL91 Q9dl91 human immun
43 58 19.1 1951 2 Q7P2K8 Q7p2k8 fusobacteri
44 58 19.1 3118 2 Q17575 Q17575 caenorhabdi
45 57.5 18.9 129 2 Q8TKY8 Q8tky8 methanosarc

ALIGNMENTS

RESULT 1
Q48496 PRELIMINARY; PRT; 81 AA.
AC Q48496
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Acidocin A precursor.
GN Name-acda;
OS Lactobacillus acidophilus.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1579;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95314239; PubMed=7793908;
RA Kanatani K, Ohnura M., Sano K.;
RT Isolation and characterization of acidocin A and cloning of the
RT bacteriocin gene from Lactobacillus acidophilus.";
RL Appl. Environ. Microbiol. 61:1061-1067(1995)
DR EMBL: D37881; GenBank: U07120.1;
DR InterPro; IPRO10133; Bacteriocin_sig.
DR TIGRFAMS; TIGR01847; Bacteriocin_sig; 1.
KW SIGNAL.
FT CHAIN 1 23 mature acidocin A.
SQ SEQUENCE 81 AA; 8946 MW; 89698AA296F7819D CRC64;
Query Match 64.1%; Score 195; DB 2; Length 81;
Best Local Similarity 64.6%; Pred. No. 1.1e-17;
Matches 42; Conservative 1; Mismatches 4; Indels 18; Gaps 3;

QY 1 KTYGTNGVHCTKNSLWGVKRLKNMKYDNTYMGRL-----QD--ILLGWATGAF 49
DB 24 KTYGTNGVHCTKNSLWGVKRLKNV-----IPGTLCKRQSLPIKQDLKILLGWATGAF 76
QY 50 GKTFH 54
DB 77 GKTFH 81

RESULT 2

GLNA_SULSO STANDARD; PRT; 471 AA.
ID GLNA_SULSO
AC P23794;
DT 01-NOV-1991 (Rel. 20, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Glutamine synthetase (EC 6.3.1.2) (Glutamate--ammonia ligase) (GS).
GN Name-gltna; Synonyms=gltna-1; OrderedLocNames=SSO0366;
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Thermoprotei; Sulfolobales; Sulfolobaceae;
OC Sulfolobus.
OX NCBI_TaxID=2287;

RP SEQUENCE FROM N.A.
RX MEDLINE=90318316; PubMed=1973523;
RA Sanangelantoni A.M., Barbarini D., di Pasquale G., Cammarano P.,
Tiboni O.;
RT "Cloning and nucleotide sequence of an archaeobacterial glutamine
synthetase gene: phylogenetic implications.";
RL Mol. Gen. Genet. 221:187-194(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=21332296; PubMed=11427726; DOI=10.1073/pnas.114222098;
RA She O., Singh R.K., Confalonieri F., Zivanovic V., Allard G.,
Awaysz M.J., Chan-Weher C.C.-Y., Clausen I.G., Curtis B.A.,
De Moore A., Erauso G., Fletcher C., Gordon P.M.K.,
Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon Sulfolobus solfataricus P2";
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate +
L-glutamine.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the glutamine synthetase family.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; X53263; CA37353.1; -.
DR EMBL; X53263; CA37352.1; -.
DR EMBL; AB006669; AAK40696.1; -.
DR PIR; A99180; A99180.
DR PIR; S11899; S11899.
DR HSSP; P06201; ILGR.
DR InterPro; IPR004809; GlnA.
DR InterPro; IPR008147; Gln_synth_beta.
DR InterPro; IPR008146; Gln_synth_C.
DR Pfam; PF00120; Gln-synth_C; 1.
DR Pfam; PF03951; Gln-synth_N; 1.
DR ProDom; PD001057; Gln_synth_C; 1.
DR TIGRFAMs; TIGR00653; GlnA; 1.
DR PROSITE; PS00180; GlnA_1; 1.
DR PROSITE; PS00181; GlnA_ATP; 1.
KW Complete proteome; Ligase.
FT BINDING 400 400 AMP (under conditions of abundant
FT glutamine) (by similarity).
FT S -> T (in Ref. 1).
FT
SQ SEQUENCE 471 AA; 53347 MW; 1CC4239287B6C2CD CRC64;

Query Match 21.5%; Score 65.5; DB 1; Length 471;
Best Local Similarity 48.6%; Pred. No. 7.7;
Matches 17; Conservative 2; Mismatches 13; Indels 3; Gaps 2;

QY 1 KTYVGTN--GVHCTKNSLWGVRLKNMKYDQNTTY 33
DB 255 KPIYDNGTGMH--THLSLWTKDGKKNLMYDPNDEY 288

RESULT 3
Q75CA6 PRELIMINARY; PRT; 1071 AA.
AC Q75CA6;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE ACR006Cp.
GN ORFNames=ACR006C;

OS Ashbya gossypii (Yeast) (Eremothecium gossypii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Eremothecium.
OX NCBI_TaxID=33169;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 10895;
RA Brachat S., Voegelé S.E., Dietrich F.S., Lerch A., Gaffney T.,
Philippsen P.;
RL Submitted (SEP-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AS016816; AAS51233.1; -.
DR AGD; ACR006C; -.
SQ SEQUENCE 1071 AA; 123572 MW; 0816B52D60032663 CRC64;

Query Match 21.2%; Score 64.5; DB 2; Length 1071;
Best Local Similarity 31.7%; Pred. No. 26;
Matches 13; Conservative 8; Mismatches 11; Indels 9; Gaps 1;

QY 8 GVHCTKNSLWGVRLKNMKYDQNTTYMGRQLDILLGWATGA 48
DB 702 GTAMAKNSLYGCLSVHNIXEENAWYL-----WITSS 733

RESULT 4
SASP_LACSK STANDARD; PRT; 61 AA.
ID SASP_LACSK
AC F35618; Q57121;
DT 01-JUN-1994 (Rel. 29, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Bacteriocin sakacin P precursor (Sakacin 674).
GN Name=sakP; Synonyms=sakR, sppA;
OS Lactobacillus sakei.
OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;
OC Lactobacillus.
OX NCBI_TaxID=1599;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Lb673;
RX MEDLINE=94236240; PubMed=8180701;
RA Tichaczek P.S., Vogel R.F., Hammes W.P.;
RT "Cloning and sequencing of sakP encoding sakacin P, the bacteriocin
RT produced by Lactobacillus sake LTH 673";
RL Microbiology 140:361-367(1994).
RN [2]
RP SEQUENCE FROM N.A.; SEQUENCE OF 19-61, AND CHARACTERIZATION.
RC STRAIN=Lb674;
RX MEDLINE=94186010; PubMed=8138128; DOI=10.1016/0378-1097(94)90005-1;
RA Holck A.L., Axelsson L., Huehne K., Kroeckel L.;
RT "Purification and cloning of sakacin 674, a bacteriocin from
RT Lactobacillus sake Lb674";
RL FEMS Microbiol. Lett. 115:143-150(1994).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=Lb674;
RX MEDLINE=96262715; PubMed=8704983;
RA Huehne K., Axelsson L., Holck A., Kroeckel L.;
RT "Analysis of the sakacin P gene cluster from Lactobacillus sake Lb674
RT and its expression in sakacin-negative Lb. sake strains.";
RL Microbiology 142:1437-1448(1996).
RN [4]
RP SEQUENCE OF 19-59, AND CHARACTERIZATION.
RC STRAIN=Lb673;
RA Tichaczek P.S., Nissen-Meyer J., Nes I.F., Vogel R.F., Hammes W.P.;
RT "Characterization of the bacteriocins curvacin A from Lactobacillus
RT curvatus LTH1174 and sakacin P from L. sake LTH673";
RL Syst. Appl. Microbiol. 15:460-465(1992).
CC -1- FUNCTION: Bacteriocidal activity; inhibits closely related
CC Lactobacilli, Listeria monocytogenes and ivanovii, Enterococcus
CC faecalis, Carnobacterium sp and Brochothrix thermosphacta.
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- SIMILARITY: Belongs to the bacteriocin class IIA/YNGV family.
CC

[illegible]

```

KW Complete proteome.
SQ SEQUENCE 1630 AA; 172741 MW; 38E9CE42F037B75A CRC64;

Query Match      20.7%; Score 63; DB 2; Length 1630;
Best Local Similarity 35.2%; Pred.No. 64;
Matches 19; Conservative 6; Mismatches 21; Indels 8; Gaps 3;

Qy 4 YTGNGVHCTKNSLWGKVRLKNMYDQNTYMGRLQDILL---GWATGAFGKTFF 53
   :|||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 1358 FGTINGEYKTDTA--GVIDYKNHAY--GVAYVHENEDIKLGRGIGWYTGIVHTPF 1407

RESULT 8
GLNA_METVO STANDARD; PRT; 446 AA.
AC A P21154;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DE 03-JUL-2004 (Rel. 44, Last annotation update)
DE' Glutamine synthetase (EC 6.3.1.2) (Glutamylammonia ligase) (GS).
GN NamesglN;
OC Methanococcus voltae.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanococcaceae; Methanococcus.
OX NCBI_TaxID=2188;
RN [1]
RZ SEQUENCE FROM N.A.
RC STRAIN=DSM 1537 / PS;
RX MEDLINE=90119872; PubMed=2575777; DOI=10.1016/0923-2508(89)90012-0;
RA Possett O., Sibold L., Aubert J.-P.;
RT "Nucleotide sequence and expression of the glutamine synthetase structural gene, glnA, of the archaeobacterium Methanococcus voltae.";
RL Res. Microbiol. 140:359-371(1989).
CC -1- CATALYTIC ACTIVITY: ATP + L-glutamate + NH(3) = ADP + phosphate + L-glutamine.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: Belongs to the glutamine synthetase family.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.eib-sib.ch/announcements or send an email to license@sib-sib.ch).
-----
DR ENBL; X53509; CAA37585.1; -.
DR PIR; A43995; A43995.
DR HSP; P06201; ILGR.
DR InterPro; IPR008147; Gln_synth_beta.
DR InterPro; IPR008146; Gln_synth_C.
DR InterPro; IPR004809; GlnA.
DR Pfam; PF00120; Gln-synth_C_1.
DR Pfam; PF03951; Gln-synth_N_1.
DR ProDom; PD001057; Gln_synth_C_1.
DR TIGRFAMs; TIGR00653; GlnA_1.
DR PROSITE; PS00180; GLNA_1.
DR PROSITE; PS00181; GLNA_ATP_1.
DR KW Ligase.
SQ SEQUENCE 446 AA; 50199 MW; 95449E4DE8542690 CRC64;

Query Match      20.6%; Score 62.5; DB 1; Length 446;
Best Local Similarity 33.9%; Pred.No. 18;
Matches 19; Conservative 9; Mismatches 17; Indels 11; Gaps 5;

Qy 1 KTYYGTVN--GVHCPTKSLW--GKVRLLKNMKYDQNTYMGRLQDILLGWATGAFGKT 52
   |:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 237 KPFFGMNGSGWHCNQ-SIWLGRKPSF----YDENNAH--QLSDICLSYIGILEHT 285

RESULT 9
Q8VM63 PRELIMINARY; PRT; 558 AA.
ID Q8VM63

```

```

Q9DL93
ID Q9DL93 PRELIMINARY; PRT; 853 AA.
AC Q9DL93;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Envelope glycoprotein.
GN Name=env;
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20541986; PubMed=11090186;
RY DOI=10.1128/JVI.74.24.11858-11872.2000;
RU Hu, Qx, Barry A.P., Wang, Zx, Connolly S.M., Peiper S.C.,
RA Greenberg M.L.;
RT "Evolution of the human immunodeficiency virus type 1 envelope during
infection reveals molecular correlates of specificity for coreceptor
utilization and AIDS pathogenesis.";
RL J. Virol. 74:11858-11872(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Hu Q.-X., Barry A.P., Wang Z.-X., Connolly S.M., Goodman D.,
RA Peiper S.C., Greenberg M.L.;
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF310113; AAG36991.1; -
DR HSSP; P04578; 1DLB
DR GO; GO:0016021; C:integral to membrane; IEA.
DR GO; GO:0019028; C:viral capsid; IEA.
DR GO; GO:0019031; C:viral envelope; IEA.
DR GO; GO:0005198; F:structural molecule activity; IEA.
DR InterPro; IPR000328; Env GP41.
DR InterPro; IPR000777; GP120.
DR Pfam; PF00516; GP120; 1.
DR Pfam; PF00517; GP41; 1.
DR KW; AIDS; Coat protein; Envelope protein; Glycoprotein; Polyprotein;
KW Transmembrane.
SQ SEQUENCE 853 AA; 96901 MW; 4BB4C83101B632E7 CRC64;

Query Match 20.2%; Score 61.5; DB 2; Length 853;
Best Local Similarity 35.9%; Pred. No. 50;
Matches 14; Conservative 8; Mismatches 14; Indels 3; Gaps

QY 7 NGVHCTKSLGWKVR--LKNMKYDQNTYMGRLQDILL 42
DB 136 NDTSGTSSWGKVGGEIKNCFSNFTTSIRGVQEYSL 174

RESULT 14
Q9DL79 PRELIMINARY; PRT; 858 AA.
AC Q9DL79;
DT 01-MAR-2001 (T-EMBLrel. 16, Created)
DT 01-MAR-2001 (T-EMBLrel. 16, Last sequence update)
DT 01-JUN-2003 (T-EMBLrel. 24, Last annotation update)
DE Envelope glycoprotein.
GN Name=env;
OS Human immunodeficiency virus 1.
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.
OX NCBI_TaxID=11676;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20541986; PubMed=11090186;
RY DOI=10.1128/JVI.74.24.11858-11872.2000;
RU Hu, Qx, Barry A.P., Wang, Zx, Connolly S.M., Peiper S.C.,
RA Greenberg M.L.;
RT "Evolution of the human immunodeficiency virus type 1 envelope during
infection reveals molecular correlates of specificity for coreceptor
utilization and AIDS pathogenesis.";
RL J. Virol. 74:11858-11872(2000).
RN [2]
RP SEQUENCE FROM N.A.

```

